A Semiotic Framework for the Semantics of Digital Multimedia Learning Objects

May, Michael

Published in:
14th International Conference on Image Analysis and Processing Workshops, 2007. ICIAPW 2007.

Link to article, DOI:
10.1109/ICIAPW.2007.8

Publication date:
2007

Document Version
Publisher's PDF, also known as Version of record

A Semiotic Framework for the Semantics of Digital Multimedia Learning Objects

Michael May
LearningLab DTU, Technical University of Denmark
mma@dtv.dk

Abstract

The relevance of semiotics for extending multimedia description schemes will be shown relative to in existing strategies for indexing and retrieval. The semiotic framework presented is intended to support a compositional semantics of flexible digital multimedia objects. Besides semiotics insights from Formal Concept Analysis is utilized.

1. Introduction

The main focus of current discussions about *reuse and flexibility of multimedia content* have been on the personalization of content in the context of leisure activities, e.g. entertainment. The attempt in the present paper of introducing a *semiotic framework* for the description of flexible multimedia content will initially require a shift in focus towards "regulated" domains and "professional" activities where digital multimedia objects and documents are being reused and tailored for specific purposes within e.g. engineering education, simulator training, or within specific work activities such as supervisory control in process plants. It is assumed however that the barriers between classical contexts-of-use (e.g. entertainment, education, training, and work) will tend to break down in the future with regard to the reuse and flexible deployment of multimedia content. A key issue is here the design of *digital learning object repositories,* because flexible component-based digital objects are well suited to play different roles across different context: a composite display required for supervisory control might contain multimedia content and Graphical User Interface components (GUIs) that would be well suited for training within another context, as well as content that could reappear in new combinations as "documents" and "measurements" to support maintenance work in yet another context. To support this *extended reuse and flexibility across contexts* (even within restricted and well-defined domains and activities), we have to provide a *compositional semantics* for multimedia content as it is transformed through recombination of media elements, "transcoded" between media types, transformed in its expressive use of different sign types, or just regrouped in its spatial or temporal layout to adapt to different devices or user preferences. Current standards for multimedia content description (e.g. MPEG-7) and formats like SMIL and SVG are insufficient because they confuse different aspects of compositionality [16], i.e. the *selection* and *composition* of media elements (as well as the separate issue of representational forms) and the spatial and temporal layout of assembled media elements.

2. Two Semantic Gaps

From the point of view of semiotics (the theory of signs and signification [6]) two research problems in multimedia theory and the theory of digital libraries are intimately related: *multimedia content description* for indexing and retrieval of composite multimedia objects and *metadata descriptions of learning objects* for digital repositories share a series of semantic problems that could benefit from a semiotic analysis. Learning objects can be considered as coextensive with digital objects except for the educational *use context* of learning objects and the associated requirement of *didactic metadata* to describe this intended use.

Initially the problem of the so called *semantic gap* in indexing and retrieval can be identified as a common problem for the description of multimedia databases, for composite multimedia on the semantic web, and for modular multimedia learning objects. It can be shown however that there is a *secondary* semantic gap, namely the semantic gap arising from the *contextual* nature of the articulation and understanding of meaning.

The *meaning* of any multimedia content in actual use cannot be fully determined on the *lexical level* ("words") of objects and events represented within the given *content*, nor can it be determined on the level of the underlying "low-level" *features* of its physical *presentation* alone. The relation between these two levels, the feature-based and the lexical semantics, constitutes the *primary semantic gap*.

14th International Conference of Image Analysis and Processing - Workshops (ICIAPW 2007)
0-7695-2921-6/07 $25.00 © 2007 IEEE

The existing strategies for indexing and retrieval of digital image, video, audio or multimedia objects have mainly focused on the top-down approach of providing *user-centred descriptions of represented objects and events*, and on the bottom-up approach of extracting *low-level graphical* (e.g. color, texture, shape, motion) or *acoustic* (e.g. pitch, timbre) *features*. Integrated approaches rely on combinations of top-down and bottom-up strategies. A significant novelty was the introduction of *relevance feedback* as a way to incorporate the semantic intuition of users in retrieval by using the semantic discrepancy of search results based on low-level features as a filter to narrow down further search results [15]. A related approach is the systematic attempt to *maintain and reuse the emergent semantics* arising from authoring and personalization of multimedia databases such as photo collections [16]. Another significant contribution was the insight that the semantic gap can be reduced, if we construct a richer semantics based on the combination of elements into *phrases*, i.e. the next and more expressive level above the "word" level in the *hierarchy of signification* [2].

A *secondary semantic gap* however arises from the *specification of meaning on these levels "above" the lexical level*, i.e. on a *phrastic* level ("sentences"), a *narrative* level ("story" and "plot"), *discursive* level ("rhetoric") and on a *pragmatic* level ("interaction") [12]. It is not just a matter of levels, but also a matter of the *specification of multimedia content within different contexts-of-use*. As long as we stay on the level of lexical meaning and/or (low-level) features, the contextual nature of meaning, i.e. when multimedia content is actually articulated and understood, does not impose itself upon us. It is when we analyze multimedia content in actual use situations that we discover the necessity of specifying meaning on more advanced levels such as the *narrative structures* of "story" and "plot" in fictional movies, educational video sequences, computer games, TV series, or sports videos, or such as the *discursive structures* embedding arguments and distributing rhetorical roles in interactive learning objects. The contextual nature of meaning is sometimes understood as if there can be no compositional semantics possible for multimedia, but this is a mistake since we still need semantics on the level of features and "words" in order to make sense of situated phrases, i.e. as in the case of natural language. We will return to the second gap after reconsidering the primary gap.

2. The Primary Semantic Gap

The primary semantic gap is a consequence of the structural aspects of language and the positional or "differential" nature meaning in general. The meaning of a text, an image, an audio sequence, a video sequence, or a multimedia object cannot entirely be *anchored* in the physical properties of its media of presentation. Some meaning is *expressed* through the physical media (e.g. graphics, acoustics) but the meaning will never the less be partially detached from its *substance of expression* (using a concept of L. Hjelmslev). The meaning of an image or a video sequence is never reducible to the visual or auditory recognition of objects and scenes represented within it.

An exemplification of this "non-materiality" of signs is the case where the interpretation of an image or a video sequence is dependant on *an object that is not present* in the depicted scene (e.g. a soccer player that is missing on the field in a particular part of a match), or dependant on *an action that is absent* from an event (e.g. a goal in a soccer match that was an obvious opportunity, but not accomplished by the players). The attempt to identify events and actions by their low-levels features such as observable *elementary motion units* [7] will fail in these cases, because the meaningful units here are *absent actions* (i.e. potential actions, but not accomplished). Even though they are absent they can still have real effects (e.g. not hitting the ball at the right moment, not reading a love letter). The objects and events "missing" from the visual scenes will on the other hand often be the focus of commentaries added by other representational forms, especially in the form of natural language (e.g. sports commentaries), since this is how we share information about objects or events that are not part of a present observable situation in which we participate (e.g. as spectators).

A *type of action* (e.g. scoring a goal in a soccer match) cannot in general be identified in advance with particular *movement descriptors* even if a material anchoring is given, because the action as a type refers to an *equivalence class* of events which will have the same semantic description, i.e. the same meaning. Since this description will abstract from the particular manner in which the action was carried out, there is not necessarily any unity to find at the level of movement description (although we can sometimes retrospectively infer a particular action by reasoning backwards from its observable result).

A similar conclusion was reached in a key paper on semantics in visual information retrieval [2]: a word can stand for multiple images, because it represents an *equivalence class* of objects, "thus reflecting a higher semantic level than that of the objects themselves". This brings us again to the very question of *levels of meaning* that is addressed in the present paper. The

cited paper ends with the insight that future multimedia retrieval systems "will have to support access to information at different semantic levels to reflect diverse application needs and user queries" [2].

3. Content Form and Expression Form

The problem of the limited anchoring of image meanings in their physical *substance of expression* can be understood through the semiotic layering of *form* and *substance* within an *expression* as separated from its *content*. Any system of signs should be understood as a language with a separation of *expression* and *content* in the signs and constructions of the language (mere *signals* do not imply such a separation), but expression and content each have a form which is relatively independent of its substance (fig. 1), cf. the use of the model to specify syntax and semantics of *multimedia units* in [13].

The *forms of expression* for multimedia digital objects are the *abstract sign types* (e.g. Image, Map, Graph, Diagram, Network chart, Language, Symbol) and their articulation within media as *representational forms* (graphical images, acoustic images, graphic maps, acoustic maps etc.) [10], as well as the *higher levels of articulation of meaning* emerging as a result of selection, combination and modification of these simple forms into composite multimedia and multi-representational objects. The higher levels or the "*hierarchy of signification*" [2] includes *phrastic* structures (sentence meaning), *narrative* structures (story and plot meaning), *discursive* structures (rhetorical meaning), and *pragmatic* structures (interaction and social meaning) [12].

The *substances of expression* are the physical *media* (e.g. graphical, acoustic) in which these forms are expressed and materially anchored. The *content forms* are the abstracted recurrent forms of *conceptual structures* through which we articulate and understand different domains; structures organized and represented by *ontologies* at different levels of detail, i.e. from top-level ontologies to domain and task ontologies [9].

The *substances of content* are the thoughts, ideas, perceptions and emotions realized and communicated, or in a narrower sense the *specific information content*. This content is however not completely "contained" within the digital objects, but is the information content *constructed* by human actors in specific *context-of-use*, i.e. a construction dependant on the knowledge and understanding of embodied and situated human actors within some purposeful activity.

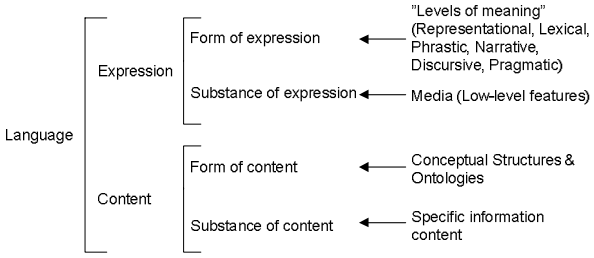

Fig.1 Aspects of digital multimedia objects.

In this semiotic model (fig. 1) the *primary semantic gap* is constituted by the fundamental separation of form and substance of signs, whereas the *secondary semantic gap* is constituted by the many *levels of signification* opened up by the form of expression, i.e. by the complex organization of meaning in language. In the case of video the levels above the object representation of the content refers to *cinematic codes* for montage, narrative sequencing etc. [3].

4. Iconicity and Representational Forms

The contextual nature of meaning does however not imply that the meaning of images, videos and multimedia objects are completely conventional or that "the meaning of the image data can only emerge from the interaction with the user" [15]. Interestingly one of the founding fathers of modern semiotics, Umberto Eco, is used to make this radical claim about image meaning, but Eco was mistaking in his early conception of *iconicity* as purely conventional, as he himself has addressed later [6]. Eco's later view gives priority to a natural iconicity based on perception, but he now seems to understate the symbolic regulation of iconic meanings in abstract-iconic forms. This "interleaved" determination of iconic and symbolic forms is an advanced aspect of the conceptions of iconicity and *diagrammatic reasoning* explored by C. S. Peirce.

Semiotics gives a foundation for the classification of signs according to different forms of iconicity ranging from the *concrete-iconic forms* of images and maps, over the *abstract-iconic forms* of graphs and diagrams, to the *symbolic forms* of symbols and languages [10][14]. These forms correspond to three underlying *similarity measures* analyzed by C. S. Peirce: the concrete-iconic forms rely on a *similarity of properties*, the abstract-iconic forms rely on a *similarity of relations*, and the symbolic forms rely on an "induced" *similarity of conceptual structures*. These types of similarity correspond to systematic differences in the interpretation of the main iconic forms (fig. 2): mages and maps are interpreted as referring to their

objects through a similarity of properties, graphs and diagrams are interpreted as referring to their objects through a similarity of relations, and languages and symbols are interpreted as referring to their objects through a (metaphorical) similarity of conceptual structures.

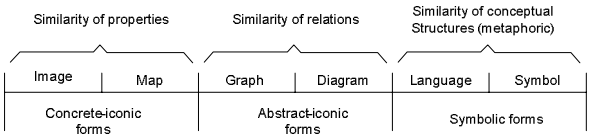

Fig. 2 Main forms of iconicity.

With few exceptions, e.g. the semiotic models of multimedia in [13][14], the classification schemes inherited from graphical design have been incomplete (in only considering types of digital objects based in established data types) as well as inconsistent (in confusing sign types and media types). The relevance of *abstract sign types* relies in their *core semantics* independent of the specific content they are used to convey and this "minimal core" can even be defined as *invariant* across different media of presentation (e.g. graphic, acoustic) [10]. The *semantic features* of sign types can be explored by Formal Concept Analysis (FCA) [8] and used to construct a formal *feature-based classification* represented as lattice structures.

The feature structure approach to sign types is basically the construction of a lattice of the logical combinations of the semantic features claimed for the types. In FCA the features (called "attributes") and the concepts (called "objects") they specify are related according to a matrix called *a formal context*. A formal context C:= (G, M, I) is defined as two sets G (from German "Gegenstände", Objects) and M (from German "Merkmale", attributes) with a relation I between G and M. The elements of G are the objects and the elements of M are the attributes or features of the context. From the formal context, all possible combinations of formal concepts can be generated. A *formal concept* is a pair (A, B) where A is a subset of the set of objects G, B is a subset of the set of attributes M, and where A´=B and B´=A (i.e. AxB is a maximal subset of I). Lattices [11] are well-suited to express feature structures and conceptual structures, because they can express inheritance relations as well as the systematic combination of types. We can use lattices to *generate possible combinations* of sign types and media types in cases where we might not know examples in advance, i.e. we can use lattices to interactively *explore the design space* of all possible combinations and their expression, and FCA have also been used for direct *browsing of image databases* [5].

Attributes for a *formal context* of sign types can be constructed from an analysis of the differential properties of examples of the media-specific forms within different content domains, but it is essential to understand that empirical forms are usually multi-representational presentations with many layers of meaning. The attributes will not be explained here, see [11], but a set of features that have been used for exploratory analysis is shown below (fig. 3) together with an example (fig. 4) of a *sub-lattice* of selected types from the resulting lattice. The features (attributes) are shown in the upper part of the lattice and the sign types are shown just above the bottom element. From features and sign types (objects) we can construct a matrix showing the assumed conceptual relations holding for this formal context.

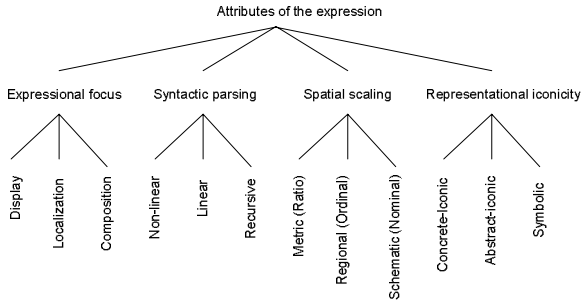

Fig. 3 Attributes of the expression used in the exploration of the formal context of sign types

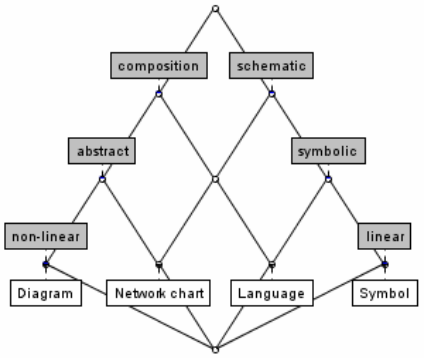

Fig. 4 Sub-lattice of a few sign types.

The unit *representational forms* (graphic image, acoustic image etc.) can now be defined through the *product lattice* of the lattice of abstract sign types and the lattice of media types (constructed from the feature analysis of media types, i.e. graphic, acoustic, gestic, haptic). Given an application domain with top-level ontology, and with domain and task ontologies, the method of Representation Design will be to utilize this generic knowledge about media types and sign types to select, construct and modify adequate combinations of media-specific representational forms to match the *content forms* required by the domain and the task.

0-7695-2921-6/07 $25.00 © 2007 IEEE

5. Flexible information presentation

An example of this abstract matching of sign types to content forms is given below from the work domain of supervisory control, where the top-level ontology has identified the relevant *content forms* for measurement data and data extracted from documents and databases as either having the form of (1) *variables* (e.g. pressure, temperature), (2) *constraints*, i.e. relations between variables (e.g. flow, temperature history), and (3) *objects and object-relations* (physical components, causal relations between alarms). In fig. 5 the match between variables and the image type can be understood as in scientific visualization theory, where a 2-D graphical image is an array of an array of data points that represent variables, whereas e.g. a (graphical) graph is a schematization in graphic space of selected relations between variables (i.e. constraints). If we need to represent data about objects (physical as well as conceptual) and object-relations, we need to shift representational form to diagrams, symbols or language.

Abstract sign type	**Type of correspondence**	**Content forms** (according to top-level ontology)
Image	*Mapping* of properties	Variables
Map	*Mapping* of constraints	Constraints
Graph	*Schematization* of constraints	
Diagram	*Schematization* of objects and object- relations	Objects and object relations
Symbol	*Categorization* of objects	
Language	*Schematization* of situations and events	

Fig. 5 Exemplification of the "match" between content forms and abstract sign types-

The purpose of specifying these abstract regularities is that we can better support *flexibility* of information presentation and interaction at the more concrete level, when we have access to a specification of the whole *design space of possible content forms and possible forms of expression* rather than just the pre-selected *specific information content* and its *actual media of presentation*. In the domain of supervisory control, work is being done on *demonstrating* the relevance of "smart instruments" and "smart documents" that have access to *models of flexible information presentation* supporting operators (e.g. in diagnosis or maintenance) by tailoring or adaptation of presentations through *"transcoding" of media* (e.g. graphic to acoustic), *transformations of representational form* (e.g. bar graphs to line graphs, from diagrams to language), *transformations of scale type* of presented data (e.g. ratio scale to ordinal scale data), *transformations in discursive perspective* on data (e.g. the part-whole composition of "mimic" process diagrams to the means-ends composition of a functional Diagram Modelling Language (DML) like MFM (fig. 6).

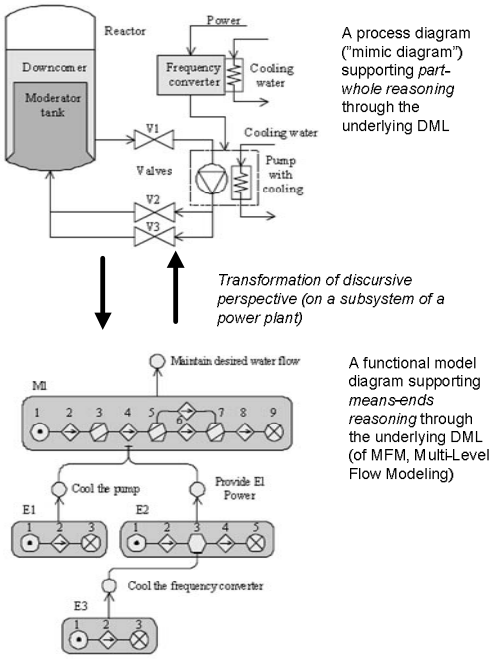

Fig. 6 A shift in discursive perspective for DML

Considered as digital multi-representational objects combining natural language text (annotation) and two sub-types of network charts ("mimic diagrams" and "MFM diagrams"), the example illustrates a more advanced form of flexibility that requires these objects to be *decomposable* into smaller *digital objects*, e.g. to support queries about a particular pump or flow function, or to support media "transcoding" or more advanced forms of transformation. Each significant part of these diagrams will have to be represented as digital objects in their own right (as XML documents) cf. SVG graphics [4], in order to support inferences within *Diagrammatic Modeling Languages* (DMLs) [1] supporting diagnostic or maintenance tasks. Different *transformations within the design space of possible media types, sign types, scale types*, as well as *transformations on higher levels of meaning*, can then be realized through XSLT-transformation on selected parts of the relevant XML documents.

The work in progress reported briefly here aims at providing the services made available by digital learning object repositories and flexible multimedia to domains characterized by well-defined ontologies and strictly regulated work practices, the example being supervisory control work. Control rooms are already *multimedia work places* blending auditory alarms,

visual displays, haptic controls etc., but HMI-design is based mainly on traditions and standards rather than on an *analytical understanding of the design space* [11]. In control rooms and distributed control work in the future we will see a more systematic utilization of *component-based multimedia services* across different devices and across different context-of-use providing an extended support for *safe, efficient and flexible information management*. To obtain this flexibility however, we have to extend multimedia description schemes beyond the *ontology-based semantic indexing* proposed for audiovisual content for leisure activities [17], in order to support "semiotic" transformations of multimedia multi-representational content. Different selections, combinations and transformations of available digital objects might reconfigure the "similar information" in the form of measurement gauges, graphs, or documents depending on the current context of use (monitoring, diagnosis, maintenance etc).

6. References

[1] Akkøk, M. N.: *Towards the Principles of Designing Diagrammatic Modeling Languages*. Ph.D. Oslo 2004.

[2] Colombo, C., Del Bimbo, A., and Pala, P. : Semantics in Visual Information Retrieval, *IEEE Multimedia* Vol. 6, July-September 1999, 38-53.

[3] Colombo, C., Del Bimbo, A., & Pala, P.: Retrieval of Commercials by Semantic Content: The Semiotic Perspective, *Multimedia Tools and Applications* Vol. 13, 2001, 93-118.

[4] Di Sciascio, E., Donini, F.M., and Mongiello, M.: A Logic for SVG Documents Query and Retrieval, *Multimedia Tools and Applications* Vol. 24(2), 125-153, 2004.

[5] Ducrou, J., Vormbrock, B. & Eklund, P.: FCA-based Browsing and Searching of a Collection of Images, *ICCS 2006, LNAI 4068*, Springer, p. 203-214.

[6] Eco, U.: *Kant and the Platypus. Essays on Language and Cognition*. Secker & Warburg, London 1999.

[7] Ekin, A., Tekalp, A. & Mehrotra, R.: Integrated Semantic-Syntactic Video Modeling for Search and Browsing, *IEEE Trans. Multimedia* 6(6), 2004, 839-851.

[8] Ganter, B. & Wille, R. : *Formal Concept Analysis. Mathematical Foundations*, Springer, Berlin, 1999.

[9] Guarino, N.: Formal Ontology and Information Systems, in: N. Guarino (ed.), Formal Ontology in Information Systems. *Proceedings of FOIS'98*, IOS Press, Amsterdam, 1998, 3-15.

[10] May, M.: Feature-based Multimedia Semantics: Representational Forms for Instructional Multimedia Design, in: Ghinea, George & Chen, Sherry Y. (Eds): *Digital Multimedia Perception and Design*. Idea Group Publishing, Hershey, 2006.

[11] May, M. and Petersen, J.: The Design Space of Information Presentation: Formal Design Space Analysis with FCA and Semiotics. *ICCS 2007 Proceedings* (Sheffield), to be published by Springer, Lecture Notes in Computer Science, 2007.

[12] May, M. *Handbook of Multimedia Semiotics and Digital Multimedia Design*, IGI Global, Hershey, to appear.

[13] Nack, F. Hardman, L.: Towards a Syntax for Multimedia Semantics, *CWI report*, Amsterdam, 2002.

[14] Purchase, H.C. and Naumann, D.: A Semiotic Model of Multimedia, in: Rahman, S.M. (ed): *Design and Management of Multimedia Information Systems*, Idea Group, Hershey, 2001, 1-21.

[15] Santini, S., Gupta, A., and Rain, R.: Emergent Semantics through Interaction in Image Databases, *IEEE Transactions on Knowledge and Data Engineering*, 13(3), 2001, p. 337-351.

[16] Scherp, A, and Boll, S.: "MM4U - A framework for creating personalized multimedia content", In: Nepal, S. and Srinivasan, U. (Eds.): *Managing Multimedia Semantics*. Idea Group Publishing, Hershey, 2005.

[17] Tsinaraki, C., Polydoros, P., Kazasis, F. and Christodoulakis, S.: Ontology-Based Semantic Indexing for MPEG-7 and TV-Anytime Audiovisual Content, *Multimedia Tools and Applications,* Vol. 26(3), 2005, 299-325.

0-7695-2921-6/07 $25.00 © 2007 IEEE